DOCUMENT RESUME

ED 064 801 EA 004 476

AUTHOR Medley, Donald M.
TITLE Supervisor Use of Observation Systems.
INSTITUTION Ohio State Dept. of Education, Columbus.
SPONS AGENCY Bureau of Elementary and Secondary Education (DHEW/OE), Washington, D.C.
PUB DATE 27 Jan 72
NOTE 10p.; Paper presented at Supervision of Instruction Symposium 2: Observation Systems and the Supervisor (January 27, 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Classroom Observation Techniques; Educational Research; *Interaction Process Analysis; Lesson Observation Criteria; *Observation; Student Teacher Relationship; Supervisors; *Teacher Behavior; *Teacher Supervision
IDENTIFIERS Elementary Secondary Education Act Title III; ESEA Title III; Observation Schedule Record; OScAR 5V

ABSTRACT

This paper discusses the supervisory use of one observational system -- Observation Schedule and Record, Form 5V (OScAR 5V). OScAR 5V refers to a series of observation schedules to develop techniques for obtaining an objective, quantitative record of observed classroom behavior that could be scored on dimensions of teacher behavior. Under this system, the observer would and should have no idea as to what the instrument is supposed to measure, his only task being to record what he sees. Taking a pupil's point of view and adopting his kind of discriminations, the observer draws only those inferences a pupil might draw about the meaning of what is said. Such inferences might include whether the teacher is telling the pupil something he should learn and remember; whether he is asking a question the pupil is supposed to answer; or whether the teacher is looking for a certain answer he has in his head. The observer records the behavior on a machine-scorable or optical scanning form. (JF)

FILMED FROM BEST AVAILABLE COPY

JUL 2 1 1972

ED 064801

Supervisor Use of Observation Systems

by

Donald M. Medley

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Presented at

Supervision of Instruction
Symposium 2:

Observation Systems and the Supervisor
January 27, 1972

Cuyahoga County Instructional Supervision Study
Richard L. Petersohn, Director
7540 Grand Division Avenue
Garfield Heights, Ohio 44125

The work presented or herein reported was performed pursuant to a grant from the Ohio Department of Education, ESEA Title III Office. However, the opinions expressed herein do not necessarily reflect the position or policy of the Ohio Department of Education, and no official endorsement by the Ohio Department of Education should be inferred.

EA 004 476

SUPERVISOR USE OF OBSERVATION SYSTEMS

Donald M. Medley - University of Virginia

The title of this paper is not entirely accurate; somewhere in the process of preparing advance materials an error was introduced which we might as well rectify now. I am really planning to talk mainly about the supervisory use of one observational system: OScAR 5V. OScAR 5V is an acronym for Observation Schedule and Record, Form 5V. To an outsider I suppose OScAR comes on like a possible alternative to Flanders' Interaction Analysis, which Dr. Amidon is discussing at this meeting, and it may be useful to look at it in that way, without (of course) becoming involved in any invidious comparisons.

We have used this acronym to refer to a series of observation schedules devised by myself and various colleagues over the past fifteen years or so, all with the same purpose. That purpose was to develop techniques for obtaining an objective, quantitativ record of observed classroom behavior which could be scored on dimensions of teacher behavior--later. The observer need--and should--have no idea of what the instrument is supposed to measure: his task is only to record what he sees.

In constructing the various versions of OScAR we have not usually known what kinds of teacher characteristics we would be measuring ourselves until some time later.

Instead, we have concentrated in each instance on developing a set of items or categories which were easy for the observer to discriminate, and which seemed to reflect as much as possible of the visible differences in teachers. This version of OScAR we are using now was five years in the making--that is, it took us about five years to define the eighteen categories which it contains. We spent those years in cut-and-try work with kinescopes of classroom teachers, which we viewed over and over, coding and recoding the behaviors on each new set of categories until we finally reached the one we have today. So far--in the last three years or so--we have not been tempted to add any new categories or revise any old ones despite some extensive field testing. So I think it is safe to share it with you as something fairly permanent.

Recording Behaviors

The observer using OScAR 5V sits in the back of a classroom during any period of time in which the teacher and all or part of the class are interacting verbally. He listens to everything the teacher says, and also to anything her pupils say that the teacher listens to. He tries to take the same point of view that a pupil would take, making the same kind of discriminations and drawing only those inferences a pupil might draw about the meaning of what is said. "Is the teacher telling me something I should learn and remember?" "Is the teacher asking me a question I am supposed to answer?" "Is he looking for some certain answer he has in his head?" and so on.

The observer records the behavior on a machine-scorable or optical scanning form (see Figure 1). Each column on the sheet

(Insert Figure 1)

provides two sets of nine usable spaces called words, and is used to describe a single event. If, for example, the history teacher says "William the Conqueror landed in 1066," he is making a statement to which no answer is expected, a statement which contains information which might be called for on a test, i.e., information the pupil sees as material he should learn and remember. A mark is therefore made in the third

space in the second word, labeled "INFAP" (Informing-Approving).

If the teacher then asks, "John, where did he land?" this is perceived as a question to which the teacher wants a specific answer, and is recorded in the next column, labeled CVG (Convergent).

The question is recorded in the next column because it begins a new event; it is recorded in the first word because it is an entry to an interchange.

There are two basic kinds of events: statements, in which only one person (almost always the teacher) is expected to speak, and interchanges, in which both a pupil and the teacher are supposed to speak, and the teacher utterance is supposed to be an evaluation of what the pupil says.

When John answers the teacher's question, the expectation is that the teacher will give him feedback about the acceptability or correctness of what he says--especially if his answer is incorrect.

John's answer is not coded, since it was not initiated by him but by the teacher's question. Only pupil initiations are coded in this system; pupil answers to teacher questions are not recorded at all.

Now suppose after John's answer, the teacher says, "That's right, in England." The teacher's response (called an exit) is recorded by marking the third space in the second word (INFAP) in the same column as the question. The entire episode--teacher question or entry, pupil answer, exit--is regarded as a single event. We have now recorded two events--a statement first, then an interchange.

And that is all there is to it. Each teacher utterance--statement, question, or exit, and each pupil initiation attended to by the teacher, is recorded by a mark in the appropriate space. With little practice the observer learns to keep pace with the normal rate of interaction, coding each utterance as it occurs. I repeat: it is neither necessary nor desirable for him to evaluate what happens or even to try to understand it. He need only code the events as they occur.

Table 1 shows the definitions of the nine categories in the first word, and Table 2 those of the nine categories in the second word. Note that a statement is recorded in

(Insert Tables 1 and 2)

terms of one mark, an interchange in terms of two. There is space on the form we now use to record 84 events in order.

A completed record presents a kind of graphic record of the behavior observed, blow by blow; and a supervisor and teacher discussing such a record immediately after the lesson took place can, with it before them, reconstruct what happened in considerable detail.

The Dataray Form

I am now going to discuss with you the use of a different form called the Dataray form which is shown in Figure 2. This form contains one space or cell for each of the

(Insert Figure 2)

60 events that it is possible to record on the form shown in Figure 1. In supervisory use, the teacher and supervisor working together can tally the events from a word to a Dataray form in a few minutes. Alternatively, the record can be "read" by a test-scoring machine and the Dataray may be printed out by a computer. In either event, a brief but informative summary of the behavior is provided; and it is the use of this form that we will discuss now.

The left-hand column of the Dataray shows teacher statements--i.e., events in which only the teacher speaks. The remainder of the form shows events in which both the teacher and a pupil speak, called interchanges. Each column shows a different kind of an interchange--that is, one beginning in a different way; and each row contains interchanges ending in a different way. Since there are eight kinds of beginnings (or entries) and seven kinds of endings (exits), there are seven times eight or 56 kinds of interchanges.

In addition, interchanges which begin with a pupil's attempt to initiate procedure (by asking permission or offering voluntarily to do something, for instance) are shown at the lower right hand corner. Since there are three kinds of endings to such events (permission given, permission refused, or neither), this brings the total number of interchanges to 59.

The interchanges on the right are pupil-initiated; those on the left, teacher initiated. Those in the lowest two rows end with negative feedback; those in the third and fourth from the top end in positive feedback.

Pupil-initiated entries are classified according to whether they deal with subject-matter or not, whether they offer to seek information, and whether they are responsive to teacher questions or not. Teacher-initiated entries are classified according to whether they are closed or open or whether they call for comment on a previous answer.

Teacher statements are classified as affective (positive or negative), procedural (directing, desisring, or describing), or as substantive--giving information or structuring problems. "Steady state" or continuing statements of the same type are also separately recorded.

In this way, although the classroom observer records behavior by coding utterances into only 18 categories, records are analyzable into 81 categories on the basis of the Dataray summary.

Eight Factor Keys

The empirical basis for Dataray interpretation is a study of secondary-school teacher. in a large metropolitan area conducted three years ago. Seventy teachers of all four academic subjects (science, English, mathematics, and social studies) were observed four times each and their behavior was recorded each time on a preliminary form of OScAR. This version permitted events to be classified into only 42 instead of 68 types. (The greater number in OScAR 5V results from subdivision of certain of the original event categories into two or more smaller subcategories).

A factor analysis of the 280 records was undertaken out of which emerged eight factors, scores on which may be said to have preserved about two-thirds of all the information contained in the 42 category frequencies, organized along eight dimensions of difference. A process of simplification led to the eight scoring keys to be described below.

Since these scores may (in a sense) be said to represent some major behavior dimensions of differences among teachers, a profile on these eight dimensions would seem to be a useful kind of summary of a record of teacher behavior. In order to get a better idea of what kind of thing OScAR measures, let us examine these scoring keys for a moment.

Four of the dimensions are related to the way a teacher deals with content, or the questioning behavior of the teacher--quantity, quality, difficulty, and source. Three

relate to the way a teacher deals with procedure or management--how much attention he gives to it, how successful he is, and how much freedom of choice he allows pupils to exercise. And one dimension has to do with the teacher's willingness to let pupils talk.

The four non-substantive keys--the ones having more to do with classroom procedu are shown in Table 3.

(Insert Table 3)

The numbers in the body of the table are scoring weights--that is, the number of behaviors of the indicated type tallied in a record is multiplied by that weight. The total of all these products (added or subtracted according to sign) is a teacher's score on that scale.

The first of the four procedural keys is called Rebuking Behavior, and is based entirely on the number of rebuking statements in a record. However, since initiating rebuking statements are weighted three times as much as continuing rebuking statements the scale may be said to reflect how often a teacher rebukes rather than how severely. (A severe rebuke tends to manifest itself in the form of a series of two or more consecutive rebuking statements--that is, it yields tallies in the continuing half of the cell.)

Managing Behavior consists essentially of a count of the number or procedural state ments--describing and directing--recorded. It also includes (with half as much weight) the difference between the number of initiating and the number of continuing supportive statements, possibly reflecting a tendency of some teachers to use considerate languag in asking pupils to do things.

Permissive Behavior contrasts the behavior of a teacher who asks a pupil what they want to do with that of one who tends to refuse permission when asked for it.

And, finally, Listening Behavior is a simple count of continuing pupil utterances--of tallies below the horizontal lines in cells 10, 20, 30, and 40. Each of these tallies represents an instance in which one pupil utterance has followed another, with the teacher remaining silent between times.

Some teachers behave like victims of a compulsion to evaluate--or at least to react to--everything a pupil says, or as though they abhorred silence in the classroom. Othe teachers appear willing to let a pupil hold the floor as long as he cares to do so--to wa after a pupil has answered a question to see whether he wishes to add anything to what has said. It is this latter type of teacher who scores on this scale, and the former who does not.

Table 4 shows the scoring keys for the four "substantive"

(Insert Table 4)

dimensions. The numbers in the body of the table are weights: to score a record each weight is multiplied by the number of behaviors recorded in the corresponding cell on th record, and the products are accumulated (with the indicated sign).

Inspection of the weights shown gives some insight into the nature of the dimensior

Lecturing Behavior is basically a count of the number of continuing informing statements, plus (with one-third the weight) the difference between the number of times a pupil gives information either voluntarily or at the solicitation of the teacher. The high a teacher's score on this dimension, the more inclined he is to tell rather than ask--in short, to lecture to the class.

Question source is a bipolar dimension contrasting teacher initiated interchanges w pupil-initiated ones. How much an interchange contributes to a teacher's score depend on the exit--that is, on how the teacher evaluates the pupil's contribution. The most

important event is a pupil-initiated interchange--substantive or non-substantive--accepted by the teacher. Each such event adds 12 points to a teacher's score! Positiv but smaller, weights also appear on pupil-initiated interchanges that are supported or neutrally rejected. Teacher-initiated interchanges with these exits on the other hand receive proportional but negative weights.

The fact that the same exits receive weights or opposite sign on pupil-initiated as compared with teacher-initiated interchanges suggests that a teacher must treat voluntary pupil comments differently than ones elicited by questions. Unsolicited comments are most likely to be accepted--that is, acknowledged but not evaluated as to correctness or acceptability--by the highscoring teacher. If positive feedback is given praise accompanies it; if negative feedback is given, it is neutral rather than critical.

When a pupil answers a teacher question, the least likely thing to happen is that the teacher will acknowledge it without evaluating it: such an event costs the teacher four points.

Since these statements describe the behavior of the teacher whose pupils initiate interchanges frequently, they seem to contain some suggestions about a teaching strategy to maximize pupil initiations. Apparently it is particularly important to accept any pupil initiation rather than to evaluate it. Apparently such behavior reinforces the pupil for saying <u>something</u>--and communicates the idea that what is rewarded in this class is participation, that you don't necessarily have to be "right" so long as you say something.

<u>Question Difficulty</u> appears at first to be a rather complex behavior pattern. Inspection reveals that the score on this key is not a function of the type (or source) of questions asked because all entries have average weights of zero. The type of exit does appear to be important, however; approving exits--especially from convergent interchanges--are the ones that count the most. Supporting, acknowledging, neutrally rejecting, and criticizing exits all tend to lower a teacher's score.

This suggests that the teacher's questions must be rather easy. Answers to easy questions tend to elicit positive feedback because they are likely to be correct, but the positive feedback does not contain support or praise as it might if the question had been a difficult one.

<u>Question Type</u> is perhaps the most complex of these four substantive keys--and the most interesting, since it seems to reflect the <u>quality</u> of a discussion, something very difficult to measure objectively. Positive weights appear almost entirely on interchanges either entered with elaborating questions or exited from without evaluation. Convergent interchanges receive a substantial negative weight if approved, accepted, or criticized.

What this suggests is that, instead of evaluating pupil's answer, the high-scoring teacher must be asking the <u>pupils</u> to evaluate or comment on it. A non-evaluating exit followed by an alaborating entry is the key behavior here, and points to a classroom in which pupils must listen to one another as well as to the teacher, because at any time they may be asked to react. And that is what a real discussion ought to be like.

Inspectional Interpretation

In supervisory uses of OScAR, the scoring keys described above may have use in measuring changes in teacher behavior over time, in developing a preliminary diagnostic profile, and possibly for evaluation. All such uses should, however, be based on records made on more than one occasion, so that behaviors peculiar to any one

situation may not be over-interpreted. Only those patterns which are stable elements of teaching style over time should be used in this fashion.

When a supervisor sits down with a teacher to discuss a lesson just observed and recorded, however, it may not be desirable to take the time to hand score the Dataray form on the factor keys. Once the behaviors have been transferred to the Dataray form by the teacher and supervisor (as described above), a simple inspection, guided perhaps by the supervisor's familiarity with the composition of the scoring keys, is probably the best next step. The supervisor may look at the form to see where the highest frequencies appear, and explain to the teacher what kind of behaviors they represent. No attempt should be made by the supervisor to evaluate at this time--he should merely describe what is there, letting the teacher become aware of the objective nature of the record. He may and should, however, ask the teacher's opinion about what the record shows. Was the teacher aware that this is what he was doing? Was this a deliberately chosen strategy for achieving the teacher's goals?

Figure 3 shows a summary for a teacher who was questioning her pupils about a

(Insert Figure 3)

story they had all read. A supervisor might point out that 57 of every 100 events recorded were non-interactive teacher statements, and that only 18 of these statements had to do with subject-matter and 39 (twice as many) had to do with managing the class.

Or he might note that of the 43 questions the teacher asked, only six were first raised for the whole class to think about before the pupil who was supposed to answer the question was identified.

Of the 28 times when bits of information were developed (shown in the informing-approving row), the teacher gave information ten times and he elicited it from the pupils 18 times.

Perhaps the most striking feature of the record is the complete absence of pupil questions or other initiations. The entire right hand side of the Dataray is blank. This can be explained--in part at least--by the large number of acknowledged and supported interchanges which the Question Source key tells us are negatively related to pupil initiations.

Problems the teacher should be able to spot are (1) a management problem possibly resulting from (2) questions that are not holding pupil interest--partly because they are too easy, and partly because they are not structured for the whole class; and (3) lack of pupil initiative.

With some guidance this teacher might decide to change her questioning behavior as a first step toward reducing her problem in management. On the next visit, then, the record should show (1) more problem structuring statements, (2) fewer approving exits, and (3) fewer directing, describing, and rebuking statements.

Figure 4 shows the record of another teacher observed in a similar situation, discussing a story the pupils have just read but following a very different strategy. This

(Insert Figure 4)

teacher uses twice as many statements (64) as interchanges (36), and practically all of them (52) deal with substantive content. Only 8 of the 100 events have to do with procedure. The teacher has asked 36 questions, 24 of which were preceded by problem structuring statements. And among the 36 questions only two elicited responses that were approved--most of the questions were either acknowledged (16) or not evaluated (14), and elaborating and divergent questions predominated.

In contrast to the last teacher whose questions were convergent and easy, this teacher is asking open questions and avoiding evaluation.

If the teacher's objective here was to elicit pupils' reactions to and ideas about the story, one would expect that, when asked how she might have done better, she might see the complete absence of pupil initiations as a problem. Knowledge of the Question Source key would lead the supervisor to suggest that the tendency to acknowledge pupil responses rather than to give feedback might be the trouble, and suggest a change here as something to try next time. Perhaps as a tactic the teacher might use procedural positive questions ("Any Questions" "Any other comments") to elicit pupil questions or statements, and then acknowledge every one she gets in order to encourage more.

Figure 5 shows a record made in a class in which the teacher was explaining new

(Insert Figure 5)

content. The striking feature here is the predominance of teacher statements--82% of the events recorded. 57 of the 82 deal with substantive content, the rest with procedure At first glance there seem to be a relatively large number of pupil initiations--as many, in fact, as there are teacher initiations. But all but two of them are pupil responses made indirectly to teacher questions.

The predominance of convergent questions approved probably indicates that the teacher questions were designed to make sure the pupils were understanding the informing statements.

Whether this is a good way to accomplish the teacher's objective is difficult to say; the teacher looking at the record and being asked how she might improve her procedures might perceive a need to get more feedback from pupils. The supervisor might suggest that the ratio of informing to describing around 4:1 shown here usually indicates a rate of information input that pupils find it difficult to follow. (An effective lecturer uses describing statements rather liberally interspersed among informing ones to do such things as pace himself, interrelate ideas, and cue pupils as to what is coming next).

Perhaps the important thing to do with this teacher is to record her behavior when she has some other objectives in mind, such as discussing homework or reviewing. If this pattern of behavior persists in such situations, some clear signs of problems may be seen.

These three examples illustrate only a minor fraction of the many interesting things in a teacher's behavior that can be documented by OScAR records. Possibly the one glaring fact is how few of the 81 events even show up at all! Teachers do not seem make use of all the ways of behaving available to them.

Notes on the Use of Observational Systems

Let me conclude my remarks by speaking for awhile to the topic assigned me; in other words let me say something about supervisors' use of observation systems in general rather than of OScAR in particular. What I am going to say represents my own inferences from research and experience in the use of observational systems not rigorously based conclusions but rather hunches or theories which you should take with a grain of salt.

As supervisors you are all charged with the responsibility of improving instruction by changing teacher behavior--that is, by helping teachers improve their teaching. This would seem to call for some definite knowledge about what good teaching is, knowledge to provide some indication of the direction in which teachers' behavior ought to change if it is to get "better." Unfortunately, we have relatively little dependable knowledge

about how a teacher ought to behave in order to be most effective. Most of what we have is sheer unsubstantiated opinion.

I am assuming here that the effective teacher is the one whose pupils show the greatest gains in achievement while they are in her class.. There is considerable research evidence to show that teachers who are rated most effective are not really the most effective in this sense of the term. I have searched the literature and every study I have found which compared supervisors' ratings or judgments of teacher effectiveness with measured gains of pupils (adjusted to equalize ability levels, etc.) has found no appreciable relationship between them. I take this as indicating that supervisors--who know more about teacher effectiveness than anyone--still do not know very much. And believe me, researchers know even less.

The practical implications of this fact seem to me to be as follows:

1. We should not be dogmatic about telling teachers how they ought to behave. Given a teacher's Dataray, Interaction Analysis Matrix, or other behavior records, I would be very careful about saying that there was too much of this or too little of that. One of my own strongest prejudices, for instance, is in favor of pupil initiations--particularly pupil statements; and if I followed my own inclinations I would probably suggest to most teachers that they should work for more of them, regardless of what their records showed. I hope, if I were a supervisor, I would be able to resist this temptation. (You will, no doubt, have noted that I yielded to it in discussing those Datarays just now).

2. We should let the teacher define for himself the changes he thinks he should make. It is a well-known principle of learning that in order to be achieved the goal of learning must be the learner's goal, not the teacher's. The learner will progrss only toward his goal, regardless of what we may be trying to teach him. Hence the importance of motivation--which is nothing more than the process of getting the learner to accept the goals we want him to reach as his own. Since we are not clear in the present instance as to what these goals should be, we would be wise to let the teacher define them for himself, and thus insure that he accepts them.

Besides, the teacher is so much closer to the problem than we are that he is probably better able to define appropriate goals than we are.

This does not mean that we should not discuss the problem and share our own experiences with the teacher in helping him define his goals, but it does mean that the final decision should be his.

3. We might as well face the fact that change itself is more important that the direction it takes. Teachers are too much inclined to get into ruts--to develop routines, habits, procedures, early in their careers and cling to them too long. I think many teachers feel helpless to change, or at least do not know how to go about it. The most important experience a supervisor can provide for a teacher is that of obtaining specific, objective, non-evaluative feedback about what is really happening while he teaches. Being able to set specific, feasible goals for change, getting concrete suggestions as to how to reach those goals, and then getting precise and objective evidence that he has indeed made the planned changes. Such an experience can and should revolutionize a teacher's attitude toward self improvement, and give him a new sense of power, of control over his own destinies to replace that helpless feeling so many teachers have.

4. Once this has taken place, improvement in instruction should be an automatic outcome. Once teachers learn how to change their behaviors at will, we should be able to leave it to them to evaluate the effects of those changes and to gradually shape their own behavior by eliminating what does not work for them, and adopting what does. Teachers actually know, better than anyone else, how effective their behavior is. They

may be reluctant to share this information with administrators who have the authority to hire, fire, and reward merit; but they have it. And they do want to improve. The supervisor can make such improvement possible--and that is all he really needs to do. Trial-and-error will do the rest.

5. Merit evaluation should be based on evidence that a teacher can change and is changing rather than on conformity of his behavior to some particular standard or model. The poorest risk is the teacher who is set in his ways--who cannot or will not change. How can he possibly adapt to the rapid changes in society's goals, methods, values, expectations that we see all about us? Inner city schools are full of teachers who were competent by yesterday's standards with yesterday's children, but who are unable to alter those once successful patterns to meet the needs of today's children. A teacher who can change will change, must change, to survive; and given time, supervisory assistance, and administrative support he will change for the better--will learn to be effective in the situation he is in.

As I said before I began this peroration, you had best look upon these suggestions as one man's opinion--strongly held. I appreciate your patience in letting me, an outsider, come here and tell you how to do your jobs.